**HADOOP 2.7.2 INSTALLING ON UBUNTU 20**

**(SINGLE-NODE CLUSTER)**

1. **Install Java 1.8.**

aramadan@ubuntu: ~$ **cd ~**

# Update the source list

aramadan@ubuntu: ~$ **sudo apt-get update**

aramadan@ubuntu: ~$ **sudo apt-get upgrade**

aramadan@ubuntu: ~$ **sudo apt-get install openjdk-8-jdk**

# Verify Java Installation

aramadan@ubuntu: ~$ **java -version**

openjdk version "1.8.0\_292"

OpenJDK Runtime Environment (build 1.8.0\_292-8u292-b10-0ubuntu1~20.04-b10)

OpenJDK 64-Bit Server VM (build 25.292-b10, mixed mode)

1. **Adding a dedicated Hadoop user.**

aramadan@ubuntu: ~$ **sudo addgroup Hadoop**

Adding group `hadoop' (GID 1003) ...

Done.

aramadan@ubuntu: ~$ **sudo adduser --ingroup hadoop hduser**

Adding user `hduser' ...

Adding new user `hduser' (1002) with group `hadoop' ...

Creating home directory `/home/hduser' ...

Copying files from `/etc/skel' ...

New password:

Retype new password:

passwd: password updated successfully

Changing the user information for hduser

Enter the new value, or press ENTER for the default

Full Name []:

Room Number []:

Work Phone []:

Home Phone []:

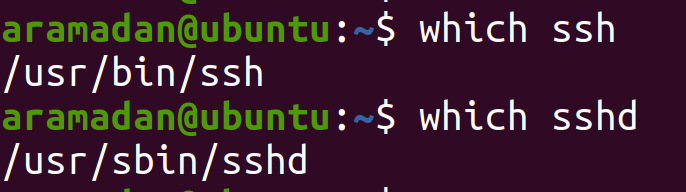
Other []:

Is the information correct? [Y/n] Y

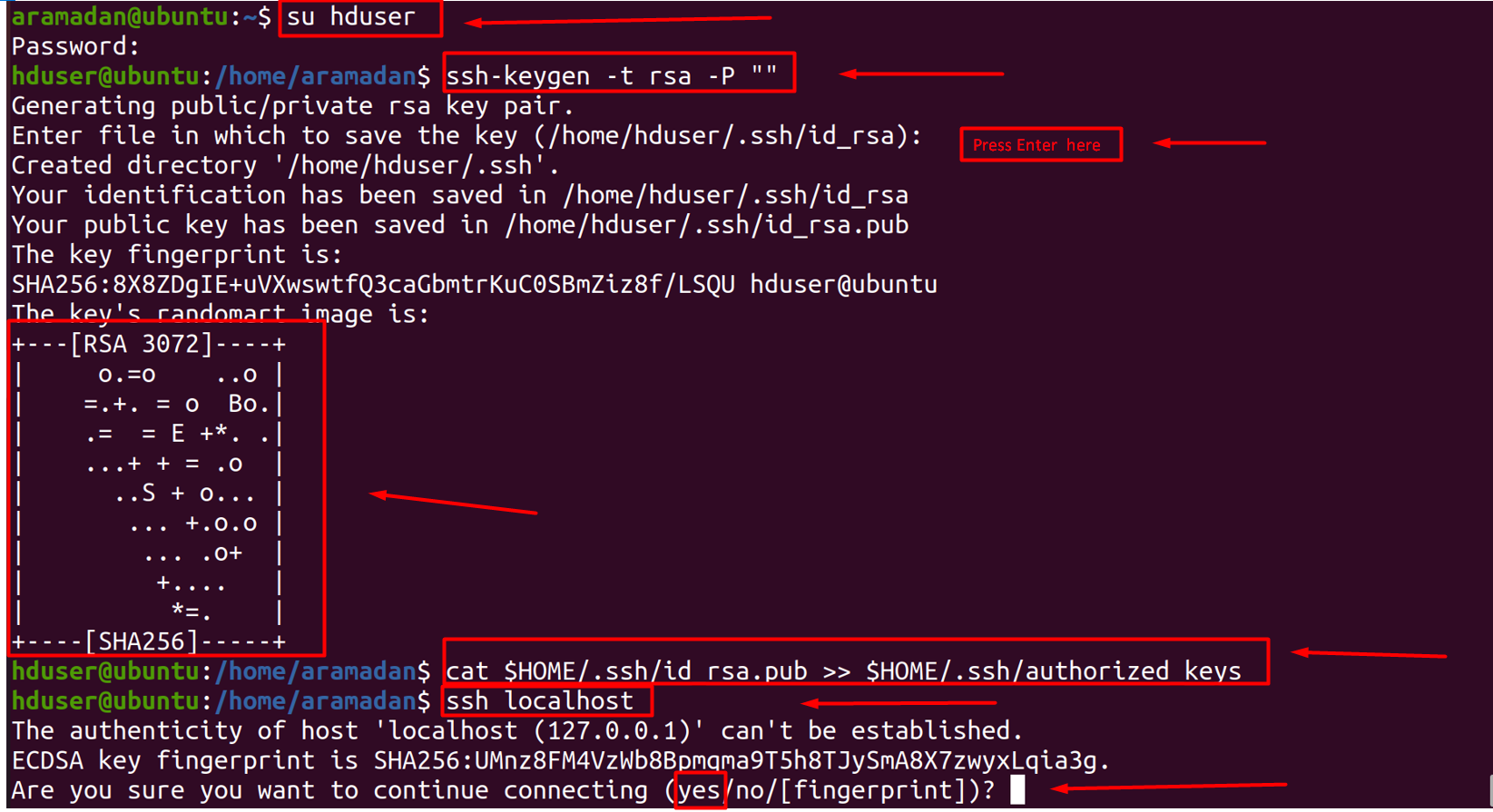
1. **Installing SSH.**
   * **ssh** has two main components:
     + **ssh** : The command we use to connect to remote machines - the client.
     + **sshd** : The daemon that is running on the server and allows clients to connect to the server.

aramadan@ubuntu: ~$ **sudo apt-get install ssh**

#Verify ssh installation



1. **Create and Setup SSH Certificates**
   * Hadoop requires **SSH access to manage its nodes**, i.e. remote machines plus our local machine. For our single-node setup of Hadoop, we therefore need to configure **SSH access to localhost**.



1. **Install Hadoop**

hduser@ubuntu: ~$ wget **"https://www.apache.org/dyn/mirrors/mirrors.cgi?action=download&filename=hadoop/common/hadoop-2.7.2/hadoop-2.7.2.tar.gz"**

#Take some time to finish the installation.

hduser@ubuntu: ~$ **tar xvzf hadoop-2.7.2.tar.gz**

hduser@ubuntu: ~$ **cd Hadoop-2.7.2**

hduser@ubuntu: ~$ **sudo mv \* /usr/local/Hadoop**

[sudo] password for hduser:

hduser is not in the sudoers file. This incident will be reported.

hduser@ubuntu: ~$ **su aramadan** #Type your primary username.

hduser@ubuntu: ~$ **sudo adduser hduser sudo** #add hduser in sudo group.

hduser@ubuntu: ~$ **sudo su hduser**

hduser@ubuntu:~/hadoop-2.7.2$ **sudo mv \* /usr/local/hadoop**

mv: target '/usr/local/hadoop' is not a directory

hduser@ubuntu:~/hadoop-2.7.2$ **sudo mkdir /usr/local/hadoop** #Create Hadoop directory.

hduser@ubuntu:~/hadoop-2.7.2$ **sudo mv \* /usr/local/hadoop**

#Verify files moving .

hduser@ubuntu:~/hadoop-2.7.2$ **ls /usr/local/hadoop/**

hduser@ubuntu:~/hadoop-2.7.2$ **sudo chown -R hduser:hadoop /usr/local/hadoop**

1. **Setup configuration Files:**
2. ~/.bashrc
3. /usr/local/hadoop/etc/hadoop/hadoop-env.sh
4. /usr/local/hadoop/etc/hadoop/core-site.xml
5. /usr/local/hadoop/etc/hadoop/mapred-site.xml.template
6. /usr/local/hadoop/etc/hadoop/hdfs-site.xml

1. **~/.bashrc**

hduser@ubuntu:~ $ **update-alternatives --config java**

**There is only one alternative in link group java (providing /usr/bin/java): /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java**

**Nothing to configure.**

hduser@ubuntu:~ $ **sudo gedit ~/.bashrc**

**# Add to the end of the file (Java & Hadoop Variables Environment)**

hduser@ubuntu:~ $ **source ~/.bashrc**

hduser@ubuntu:~ $ **javac -version**

hduser@ubuntu:~ $ **which javac**

hduser@ubuntu:~ $ **readlink -f /usr/lib/jvm/java-8-openjdk-amd64/bin/javac**

#HADOOP VARIABLES START

export JAVA\_HOME=usr/lib/jvm/java-8-openjdk-amd64

export HADOOP\_INSTALL=/usr/local/hadoop

export PATH=$PATH:$HADOOP\_INSTALL/bin

export PATH=$PATH:$HADOOP\_INSTALL/sbin

export HADOOP\_MAPRED\_HOME=$HADOOP\_INSTALL

export HADOOP\_COMMON\_HOME=$HADOOP\_INSTALL

export HADOOP\_HDFS\_HOME=$HADOOP\_INSTALL

export YARN\_HOME=$HADOOP\_INSTALL

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_INSTALL/lib/native

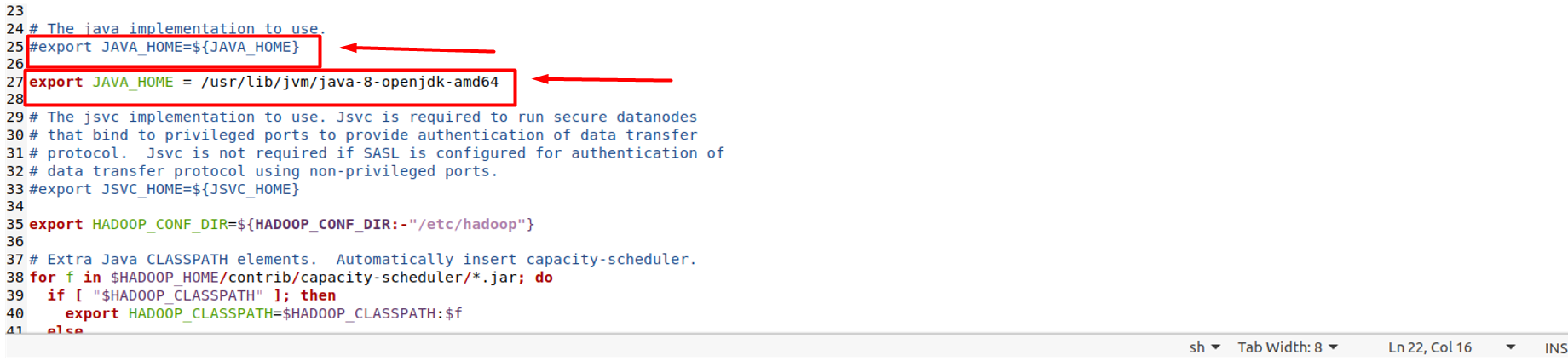
export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_INSTALL/lib"

#HADOOP VARIABLES END

1. **/usr/local/hadoop/etc/hadoop/hadoop-env.sh**

hduser@ubuntu:~ $ **sudo gedit /usr/local/hadoop/etc/hadoop/hadoop-env.sh**

**#add the following line: export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64**



1. **/usr/local/hadoop/etc/hadoop/core-site.xml**

hduser@ubuntu:~ $ **sudo mkdir -p /app/hadoop/tmp**

hduser@ubuntu:~ $ **sudo chown hduser:hadoop /app/hadoop/tmp**

hduser@ubuntu:~ $ **sudo gedit /usr/local/hadoop/etc/hadoop/core-site.xml**

**#Replace Configuration tag with the following block**

<configuration>

<property>

<name>hadoop.tmp.dir</name>

<value>/app/hadoop/tmp</value>

<description>A base for other temporary directories.</description>

</property>

<property>

<name>fs.default.name</name>

<value>hdfs://localhost:54310</value>

<description>The name of the default file system. A URI whose

scheme and authority determine the FileSystem implementation. The

uri's scheme determines the config property (fs.SCHEME.impl) naming

the FileSystem implementation class. The uri's authority is used to

determine the host, port, etc. for a filesystem.</description>

</property>

</configuration>

1. **/usr/local/hadoop/etc/hadoop/mapred-site.xml**

hduser@ubuntu:~ $ **cp /usr/local/hadoop/etc/hadoop/mapred-site.xml.template /usr/local/hadoop/etc/hadoop/mapred-site.xml**

hduser@ubuntu:~ $ **sudo gedit /usr/local/hadoop/etc/hadoop/mapred-site.xml**

**#Replace Configuration tag with the following block**

<configuration>

<property>

<name>mapred.job.tracker</name>

<value>localhost:54311</value>

<description>The host and port that the MapReduce job tracker runs

at. If "local", then jobs are run in-process as a single map

and reduce task.

</description>

</property>

</configuration>

1. **/usr/local/hadoop/etc/hadoop/****hdfs-site.xml**

hduser@ubuntu:~ $  **sudo mkdir -p /usr/local/hadoop\_store/hdfs/namenode**

hduser@ubuntu:~ $  **sudo mkdir -p /usr/local/hadoop\_store/hdfs/datanode**

hduser@ubuntu:~ $  **sudo chown -R hduser:hadoop /usr/local/hadoop\_store**

hduser@ubuntu:~ $ **sudo gedit /usr/local/hadoop/etc/hadoop/** **hdfs-site.xml**

**#Replace Configuration tag with the following block**

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

<description>Default block replication.

The actual number of replications can be specified when the file is created.

The default is used if replication is not specified in create time.

</description>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:/usr/local/hadoop\_store/hdfs/namenode</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>file:/usr/local/hadoop\_store/hdfs/datanode</value>

</property>

</configuration>

Format the New Hadoop Filesystem & Run Hadoop

hduser@ubuntu:~ $ **hadoop namenode -format**

hduser@ubuntu:~ $ **start-all.sh #Run Hadoop Services**

hduser@ubuntu:~ $ **jps**

**15888 Jps**

**15682 NodeManager**

**15218 DataNode**

**15415 SecondaryNameNode**

**15050 NameNode**

**15550 ResourceManager**